Summary: what have we learned?

Data is distributed to facilitate parallel computation

- Distributions that are good for computation are often bad for I/O
- Serial I/O interfaces are inadequate for describing multi-dimensional distributions

  Small blocks

  Lots of seeks

  Complicated seek address calculation

Summary (cont.)

I/O hardware is adequate if it is utilized well

- Disks require large contiguous reads/writes to perform well
- Software must take advantage of hardware parallelism
- I/O hardware is not free, you must buy adequate hardware to get good performance

Summary (cont.)

Collective buffering algorithms make it possible to have good compute and I/O performance

- **Must still have adequate buffer memory**
- **I/O interface must be adequate**

High level I/O interfaces such as MPI-IO can bridge the gap between UNIX style file systems and parallel applications